Data & Knowledge Engineering 68 (2009) 1359–1369

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

CBIR of spine X-ray images on inter-vertebral disc space and shape profiles using feature ranking and voting consensus

Dah-Jye Lee [a,*], Sameer Antani [b], Yuchou Chang [a], Kent Gledhill [c], L. Rodney Long [b], Paul Christensen [a]

[a] Department of Electrical and Computer Eng., Brigham Young University, Provo, UT, United States
[b] National Library of Medicine, National Institutes of Health, Bethesda, MD, United States
[c] Utah Valley Regional Medical Center, Provo, UT, United States

ARTICLE INFO

Article history:
Available online 10 July 2009

Keywords:
CBIR
Spine X-ray image
Inter-vertebral disc space
Feature ranking
Voting consensus

ABSTRACT

Very limited research is published in the literature that applies content-based image retrieval (CBIR) techniques to retrieval of digitized spine X-ray images that combines inter-vertebral disc space and vertebral shape profiles. This paper describes a novel technique for retrieving vertebra pairs that exhibit a specified disc space narrowing (DSN) and inter-vertebral disc shape. DSN is characterized using spatial and geometrical features between two adjacent vertebrae. In order to obtain the best retrieval result, all selected features are ranked and assigned a weight to indicate their importance in the computation of the final similarity measure. Using a two phase algorithm, initial retrieval results are clustered and used to construct a voting committee to retrieve vertebra pairs with the highest DSN similarity. The overall retrieval accuracy is validated by a radiologist and proves that selected features combined with voting consensus are effective for DSN-based spine X-ray image retrieval.

1. Introduction

Osteoarthritis affects a significant portion of the elderly population in the United States [1]. Osteophytes, disc space narrowing (DSN), subluxation and spondylolisthesis are typical radiographic hallmarks characterizing this condition on the spine. The ability to retrieve spine X-ray images on these conditions could be very valuable to clinicians (radiologists), researchers of arthritis and musculoskeletal diseases, and educators. This paper focuses on the problem of retrieval of digitized X-ray images of the spine based on disk space narrowing coupled with vertebral shape content analysis.

Manually finding reference images from a large image database is a tedious and error prone process. An automatic CBIR system can significantly alleviate the problem of retrieving relevant images with specified DSN. Content-Based Image Retrieval (CBIR) techniques have been studied for nearly two decades. The techniques have been used for searching images in digital libraries, on the World Wide Web, and other applications such as trademark search [2]. Research on medical image retrieval, however, has been fairly recent [3–8]. These efforts can be broadly categorized into two themes: (i) retrieval of biomedical images from a heterogeneous collection (images of different anatomy, modality, and detail) with little importance given to localized pathology, and (ii) retrieval of images from a homogenous collection (images of single modality, anatomy, and detail) with particular focus on the localized pathology. Our research [9] has been of the latter category.

* Corresponding author.
E-mail address: djlee@ee.byu.edu (D.-J. Lee).

0169-023X/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.datak.2009.07.008

Retrieval of medical images started with text-based retrieval and has grown to include image content-based retrieval with explosive growth in the acquisition and use of biomedical images. Mao and Chu [10] studied the vector space model (VSM) to automatically retrieve medical documents. Following the development of image processing and computer vision techniques, indexing and retrieval of the medical images based on content analysis became possible. Muller et al. [11] gave an overview of available literature in the field of content-based access to medical image data and on the technologies used in this field.

The Lister Hill National Center for Biomedical Communications, an intramural R&D division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), maintains an archive of digitized spine X-rays collected from the second National Health and Nutrition Examination Surveys (NHANES II) [12] which can serve as a reference collection for study of DSN. A prior study [13] proposed use of four scale-invariant, distance transform-based features to characterize spacing between adjacent vertebrae. *K*-means clustering and self-organizing map (SOM) were used to classify inter-vertebral disc space and assigned it a degree of DSN severity with an overall accuracy of 82.1%. A shortcoming of using this approach which has proven to be robust for automatically classification of severity level for shape-based CBIR is the lack of disc shape profiles. Using DSN severity level classification alone is insufficient for shape-based CBIR which is the focus of this work.

Vertebral shape is valuable in expressing the spine conditions described earlier. Fig. 1a shows two adjacent vertebrae outlined on a lumbar spine X-ray. As seen in the sagittal view, the inferior and superior edges of vertebrae adjacent to the disc can serve as the disc shape profile. Experienced radiologists use several criteria when evaluating DSN similarity between a candidate case and references from an atlas, for example. These criteria include the top to bottom size of the inter-vertebral gap, the length of the gap, and its configuration, i.e., whether there are spurs, concavities, convexities, irregularities, etc. Many of these disc space characteristics can be computed from disc shape profiles. In this paper, we present an approach that combines disc shape profile with computed inter-vertebral disc space features and uses voting consensus for finding similar images. In addressing this important problem, this effort makes advances the state of the art in CBIR taking advantage of clustering ensemble based machine learning methods [1,14–17].

The proposed algorithm and DSN similarity measures are discussed in Section 2. Feature ranking for weight computation of extracted features is presented in Section 3. Section 4 introduces the proposed voting consensus mechanism. Experimental results and analysis are presented in Section 5. Conclusions and future directions are described in Section 6.

2. DSN features selected for similarity measurement

X-ray images from the NHANES II data set used for this study are segmented using active contour segmentation method and the resulting 9-point and 36-point contour shapes are validated by a board certified radiologist. Examples of these contours are shown in Fig. 1b and c. Fig. 1b shows the 9-point model commonly used by radiologists. The left side (Points 8–7–9) of the vertebra is the anterior edge and the right side (Points 1–4) is the posterior edge. Fig. 1c shows a vertebra contour represented by 36 points. More points are used for the anterior side because it has more anatomical importance than the posterior side. In cervical and lumbar images, we use up to 4 disc inter-vertebral spaces between C3–C7 and L1–L5 vertebrae, respectively.

2.1. Mean and standard deviation of vertebra distance

In the sagittal view, a 9-point vertebral contour includes the superior and inferior "corners" on both anterior and posterior sides of the vertebra (Points 1, 8, 4, and 9). These four corners are treated as salient points to guide the calculations of spatial and geometrical features. The approximate centroid of the vertebra for feature extraction is calculated as shown in Eq. (1), where (x_c, y_c) are the coordinates of the centroid, and (x_i, y_i) are the coordinates of four corner points.

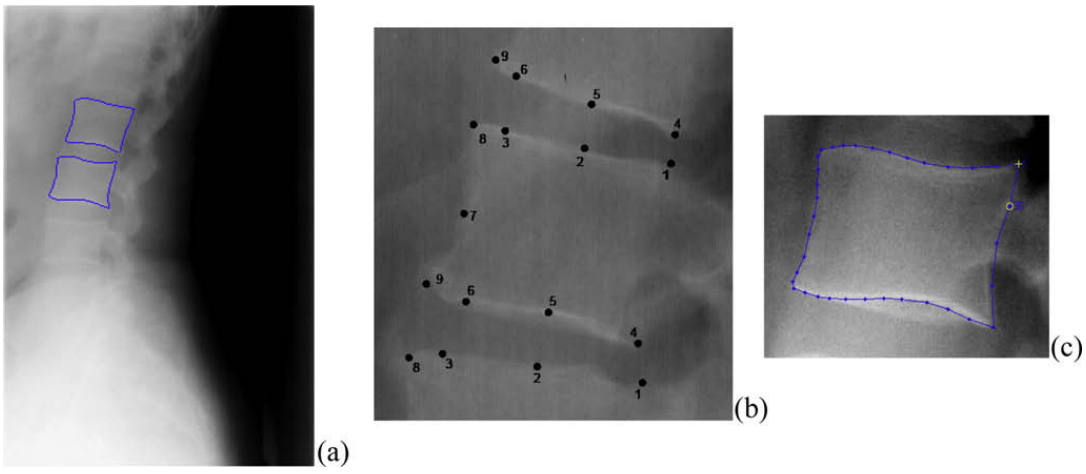

Fig. 1. (a) Spine X-ray image with the superimposed shape contours on two adjacent vertebrae, (b) 9-point model, and (c) 36-point contour.

$$(x_c, y_c) = \frac{1}{4}\sum_{i=1}^{4}(x_i, y_i) \tag{1}$$

As shown in Fig. 2, connecting the two centroids of the upper and lower vertebrae, a line segment called centroid line is formed and is divided into three segments. These sections include two segments in the interior of these vertebrae from the centroids to the edges of vertebrae and the middle segment in the inter-vertebral area (darkened line). The length of the segment in the inter-vertebral area is treated as one of many distance measures of DSN. Additionally, since the inter-vertebral disc shape profile, defined as the pair of contour edge segments between the respective inferior and superior corners of these adjacent vertebrae, are not a straight line, distance measure along the centroid line alone cannot typify the DSN. Ten line segments parallel to the centroid line that equally divide the width of the vertebra are generated. From these 11 parallel lines resulting inter-vertebral disc space distances are used to compute their mean and standard deviation and are used to represent the degree of DSN.

2.2. Vertebra skewness

The asymmetry in the disc space is computed as Skewness using polar coordinates that have previously been applied in shape representation and analysis [18,19]. It is computed as the difference between the inter-vertebral distance between anterior and posterior "corners" using the 9-point model. Two short line segments can be formed by connecting the two anterior corners and the two posterior corners. The center line, also shown in Fig. 2, can then be formed by connecting the mid points of these two short line segments. This center line is treated as the X-axis of a polar coordinate system. Furthermore, the intersection of this X-axis and the centroid line is used as the origin of the coordinate system. Based on this origin and the X-axis, the skewness of DSN is calculated as:

$$\textit{Skewness} = |(\theta_{rt} + \theta_{rb})| - |(\theta_{lt} + \theta_{lb})| \tag{2}$$

where

$$\theta = \sin^{-1}\left(\frac{\text{Distance to the } x\text{-axis}}{\text{Distance to the Origin}}\right) \tag{3}$$

In Eq. (2), θ_{rt} is the angular measurement between the posterior-inferior corner of the upper vertebra and the origin. Similarly, θ_{rb} is the angular measurement between the posterior-superior corner of the lower vertebra and the origin. Correspondingly, θ_{lt} and θ_{lb} can be computed similarly. All angular measurements are calculated using Eq. (3). The *Skewness* measure is positive when the short line segment on the right is longer than the one on the left. Otherwise, the Skewness is negative.

2.3. Normalized inertias

Besides the three features mentioned above which directly describe distance and skewness of the disk space, three shape features are also proposed as the assisting descriptors for DSN similarity measure. As shown in Fig. 1, 36-point model is interpolated to form an enclosed contour for upper and lower vertebrae, respectively. An anterior line segment is formed by connecting the anterior inferior corner (Point 9) of the upper vertebra and the anterior superior corner (Point 8) of the lower vertebra. Similarly, a posterior line segment is formed by connecting the posterior-inferior corner (Point 4) of the

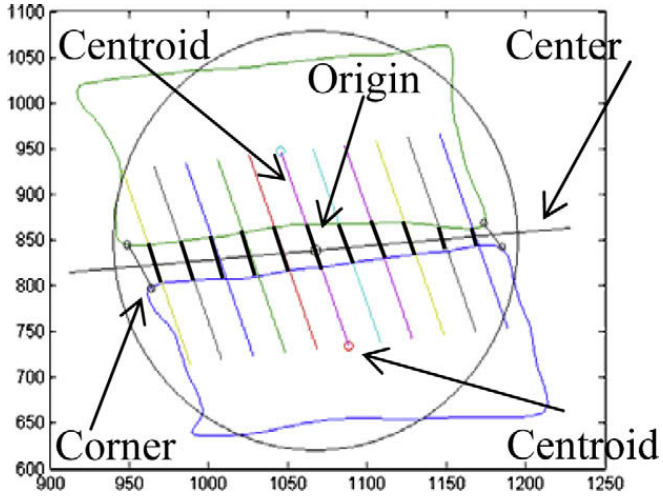

Fig. 2. Centroids of two adjacent vertebrae, centroid line that connects centroids, lines parallel to the centroid line at a fixed interval, anterior and posterior corners, and the center line that separates the vertebrae.

upper vertebra and the posterior-superior corner (Point 1) of the lower vertebra. These two short line segments (as shown in Fig. 2) along with the inferior profile of the upper vertebra and the superior profile of the lower vertebra form an enclosed contour of the disk space. One feature is needed to be calculated for this contour to describe its shape property. The normalized inertia [20] of order 2 is adopted. For the disk space region R_d in the image plane, the normalized inertia of order 2 is given as:

$$I(R_d, 2) = \frac{\sum_{r \in R_d} \|r - \hat{r}\|^2}{V_d^2} \tag{4}$$

where $\hat{r}$ is the centroid of R_d and V_d is the number of pixels in region R_d. The normalized inertia is invariant to scaling and rotation. $I(R_d,2)$ is used as the fourth feature to describe the disk space for retrieval.

Similarly, we adopt normalized inertia to describe shape property of the whole upper and lower vertebrae. Although shape property of the whole upper and lower vertebrae is not directly relevant to DSN, it serves as contextual information to retrieve similar disk spaces. Similar to Eq. (4), the normalized inertia of the upper and lower vertebrae are calculated as:

$$I(R_u, 2) = \frac{\sum_{r \in R_u} \|r - \hat{r}\|^2}{V_u^2} \tag{5}$$

and

$$I(R_l, 2) = \frac{\sum_{r \in R_l} \|r - \hat{r}\|^2}{V_l^2} \tag{6}$$

$I(R_u,2)$ and $I(R_l,2)$ represent the shape property of the upper and lower vertebrae and are used as the fifth and sixth features to describe DSN similarity.

2.4. Inter-vertebral disc shape profiles

In order to accurately retrieve the most similar vertebra pairs with similar DSN characteristics, besides the six features mentioned above, inter-vertebral disc shape profiles (the inferior profile of the upper vertebra and the superior profile of the lower vertebra) are also included. To compute profile similarity, we use the Procrustes distance measure [21]. Procrustes distance measure performs a linear transformation including translation, rotation, and scaling on one shape profile to find the best matched profiles. There is no explicit requirement that the two shape profiles be closed or open. More details about Procrustes distance measure can be found in [22].

Because Procrustes distance calculation requires an equal number of points on the two shape profiles, the first step is to localize the corresponding points on the query and the test vertebra profiles. The inferior profile of the upper vertebra is defined as the bottom curve on the upper vertebra as shown in Fig. 2. Similarly, the superior profile of the lower vertebra is the top curve on the lower vertebra. After the interpolation process on the inferior profile of the upper vertebra and the superior profile of the lower vertebra based on the 36 point model, 115 equally-spaced points on each profile are used for Procrustes distance calculation. More similar profiles result in smaller Procrustes distance.

3. Feature ranking for assigning feature weights

Feature selection has been used for assigning weights to features in CBIR in recent years [23–25]. Relevance feedback has been used to assign feature weights according to user's judgment [23,24]. Relevant features are selected and assigned greater weights [25]. Six features extracted in Section 2 are of different importance for measuring DSN similarity. They are categorized into two feature sets. Of these features, mean and standard deviation of distance, skewness, and $I(R_d,2)$ are directly related to disk space spatial and geometrical properties. The other two features, $I(R_u,2)$ and $I(R_l,2)$ are used to describe the upper and lower vertebrae. Assigning proper weights to these features according to their importance determines final retrieval accuracy.

Feature ranking [26] is a relaxed version of feature selection [27,28] which ranks all features with respect to their relevance. Feature ranking can be viewed as a kind of flexible feature subset selection approach. We use two training sets of lumbar and cervical disk spaces to measure feature relevance of the six features mentioned previously. In order to determine the importance of these features, Gaussian Mixture Modeling (GMM) [29] was adopted to use each feature to cluster the training sets into two classes (normal disk space and DSN). GMM is one of many statistically mature methods commonly used for clustering. The Expectation-Maximization (EM) algorithm [30] is used to determine the maximum likelihood parameters of a mixture of 2 Gaussians in the feature space.

The distribution of a random variable $X \in R_d$ is a mixture of k Gaussians if its density function is:

$$f(x|\theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d}} \exp\left\{-\frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)\right\} \tag{7}$$

such that the parameter set $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$ consists of:

$$\alpha_j > 0, \sum_{j=1}^{k} \alpha_j = 1$$

where $\mu_j \in R^d$ and Σ_j is a $d \times d$ positive definite matrix.

Given a set of vectors $x_1, x_2, \ldots, x_n$, the maximum likelihood estimation of θ is:

$$\theta_{ML} = \arg\max_{\theta} L(\theta|x_1, x_2, \ldots, x_n) = \arg\max_{\theta} \sum_{i=1}^{n} \log f(x_i|\theta) \tag{8}$$

EM algorithm is an iterative method to obtain θ_{ML}. Given the current estimation of the parameter set, EM algorithm estimates a new parameter set for each iteration according to the following two steps.

Expectation step:

$$\omega_{ij} = \frac{\alpha_i f(x_i|\mu_j, \Sigma_j)}{\sum_{l=1}^{k} \alpha_l f(x_i|\mu_l, \Sigma_l)} \quad j = 1, \ldots, k, \ i = 1, \ldots, n \tag{9}$$

The term ω_{ij} is the posterior probability that the feature vector x_i was sampled from the jth component of the mixture distribution.

Maximization step:

$$\begin{aligned} \hat{\alpha}_j &\leftarrow \frac{1}{n} \sum_{i=1}^{n} \omega_{ij} \\ \hat{\mu}_j &\leftarrow \frac{\sum_{i=1}^{n} \omega_{ij} x_i}{\sum_{i=1}^{n} \omega_{ij}} \\ \hat{\Sigma}_j &\leftarrow \frac{\sum_{i=1}^{n} \omega_{ij} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^T}{\sum_{i=1}^{n} \omega_{ij}} \end{aligned} \tag{10}$$

Rand Index method [31] is used to measure the accuracy of a clustering solution I. The Rand Index of the clustering solution is calculated as:

$$\|I, I^{accuracy}\| = \frac{2 \cdot (n_{00} + n_{11})}{n \cdot (n-1)} \tag{11}$$

where $I^{accuracy}$ is the ground truth of clustering solution and n is the total number of disk spaces. n_{11} is the number of pairs of disk spaces which are successfully categorized in the same DSN severity group as the ground truth. n_{00} denotes the number of pairs of disk spaces which are categorized in different DSN severity groups.

In Eq. (7), k is set to 2 to partition the 2 training sets (lumbar and cervical) into 2 classes (normal disk space and disk space with DSN). Each training set contains n = 40 disk spaces, 20 of which are normal, and the other 20 are disk spaces with osteophytes or DSN. One Rand Index (shown in Table 1) for each of the six features clustered by GMM can be calculated using Eq. (10). Features with high clustering accuracy are more important for similarity measure and are to be assigned higher weights than those with low clustering accuracy.

According to the analysis of the clustering accuracy of both feature sets (the first four features in one set and the last two features in the other set) mentioned above, the following weighting strategy is used to compute similarity measure between two disk spaces.

$$SM = \beta_1 \times D_1 + \beta_2 \times D_2 \tag{12}$$

Table 1
Clustering accuracy by GMM clustering-based on six features.

	Feature Set #1				Feature Set #2	
	Mean	SD	Skewness	DS_NI	Up_NI	Lo_NI
RI_Cervical	0.4923	0.4923	0.4872	0.5333	0.4885	0.4923
RI_Lumbar	0.4885	0.4923	0.4885	0.4987	0.4923	0.4987

Mean: mean of the distance.
SD: standard deviation of the distance.
Skewness: skew degree of the disk space.
DS_NI: normalized inertia of the disk space.
Up_NI: normalized inertia of the inferior profile of the upper disc.
Lo_NI: normalized inertia of the superior profile of the lower disc.

where β_1 and β_2 are set to 0.8 and 0.2 as weights on feature set 1 (*mean and standard deviation of distance, skewness*, $I(R_d,2)$) and set 2 ($I(R_u,2)$ and $I(R_l,2)$), respectively. β_1 and β_2 are selected to obtained the best distinction between the normal disk space and disk space with DSN. D_1 and D_2 are defined as:

$$D_1 = \alpha_{Mean}(Q_{Mean} - O_{Mean}) + \alpha_{SD}(Q_{SD} - O_{SD}) + \alpha_{Skewness}(Q_{Skewness} - O_{Skewness}) + \alpha_{DS\_NI}(Q_{DS\_NI} - O_{DS\_NI}) \quad (13)$$

and

$$D_2 = \alpha_{Up\_NI}(Q_{Up\_NI} - O_{Up\_NI}) + \alpha_{Up\_NI}(Q_{Lo\_NI} - O_{Lo\_NI}) \quad (14)$$

Q and O are six feature values for query and objective disk spaces, respectively. Coefficients α are determined according to data in Table 1.

$$\alpha_i = \frac{RI_i}{RI_{Mean} + RI_{SD} + RI_{Skewness} + RI_{DS\_NI}}, \quad i = Mean, SD, \ Skewness, DS\_NI \quad (15)$$

and

$$\alpha_j = \frac{RI_j}{RI_{Up\_NI} + RI_{Lo\_NI}}, \quad j = Up\_NI, \ Lo\_NI \quad (16)$$

In the EM algorithm for GMM distribution, among these six features, three features: mean and standard deviation of distance and skewness are directly related to describing disk space. Other three normalized inertias are related to shape properties.

4. Voting consensus

Generally, CBIR system retrieves images by comparing the query image against images in the database using similarity measures. Voting consensus has shown success in clustering ensemble [14], object classification [15], and information extraction [16]. In [17], authors used a clustering algorithm to retrieve clusters of images that are in the vicinity of the query image. These clusters can be deemed as semantic groups. This paper presents a voting consensus mechanism to achieve a similar task. Voting consensus assists in selecting the best match vertebra pairs based on multiple levels of detail and similarity measure from different perspectives. Using voting consensus for a query inter-vertebral disc, Q_0, a set of M discs with similarly expressed (using shape profiles) and with indicated proximity (using DSN measures) can be retrieved. In order to associate every retrieved image to its semantic category, K-means clustering [32] is used to partition these M retrieved images into K ($K \leqslant M$) clusters.

After the partitioning process is completed, a pre- selected number of discs, N, in the cluster that contains the original query disc, Q_0, are chosen as new queries (Q_1,Q_2,..., Q_N). Using these N new queries and the spatial and geometrical features as the similarity measures, M images most similar to each of these N new queries can, then, be retrieved.

This whole initial retrieval process provides a total of $N+1$ sets of retrieval results (the original query and N new queries) that are used as the members of the voting committee. Each of these sets of retrieval result consists of M discs most similar to its query. Furthermore, profile similarity measured using the Procrustes distance is also used to retrieve M discs to form another set of retrieval result using the original query Q_0. Adding this set of retrieval result to the voting committee results in a total of $N+2$ voting committee members. Each of the M images in their own set of retrieval result has its position on the similarity ranking list, where, the top ranked image is most similar to the query image and the image ranked at position M is least similar to the query image. The voting process is performed according to the procedure below.

(1) For each image I_i (where i represents the position of image I on the ranked response list to the original query Q_0), its vote from committee member 0 is calculated as $B_{i0} = M-(i-1)$, $i = 1,2,\ldots,M$.

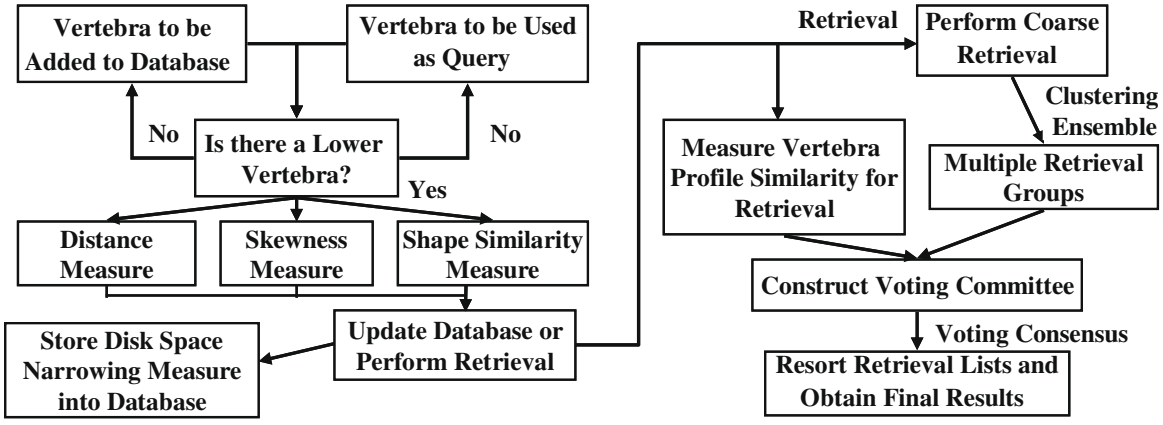

Fig. 3. Algorithm framework using six features with feature ranking and weighting.

(2) Image I_i also receives a vote from committee member j that is calculated as $B_{ij} = M - |P_0 - P_i|)$, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$. P_0 is the position of the original query Q_0 on the ranked response list to query Q_j (committee member j). P_i is the position of image I_i on the ranked response list to the original query Q_0 (committee member 0). The vote $B_{ij} = 0$ if Q_0 or I_i is not ranked as the top M in committee member j.
(3) Repeat (2) for all N sets of initial retrieval result.
(4) Image I_i also receives a vote from the last committee member that is generated by using profile similarity and the original query Q_0. The vote is calculated as $B_{i(N+1)} = M-(i-1)$, $i = 1, 2, \cdots, M$. $B_{i(N+1)} = 0$ if I_i is not ranked as the top M in the last committee member.
(5) Calculate I_i's total vote from $N + 2$ committee members from Steps 1, 2, and 4 as $B_i = \sum_{j=0}^{N+1} B_{ij}$.
(6) The final retrieval result can be obtained by sorting B_i.

This final retrieval result is then reverse ordered set of unique discs from those with most votes to those with the least. This list may be further reduced for application purposes. Fig. 3 shows the framework of the proposed algorithm using two sets of features.

5. Results

A set of 801 cervical and 972 lumbar vertebral outlines (shapes) segmented from a total of 400 digitized spine X-ray images was used for performance evaluation. Ten disc pairs of both cervical and lumbar shapes were selected randomly as queries. From the set of cervical vertebrae outlines, pairs of adjacent vertebrae were used to identify discs. Three discs from the C3–C4 pair, 2 discs from the C4–C5 pair, 3 discs from the C5–C6 pair, and 2 discs from the C6–C7 vertebrae pair were selected as queries. Five of these query shapes had osteophytes. One of them had slight osteophytes and two were moderate and two were severe. Similarly, from the set of lumbar vertebrae outlines, 3 disc pairs of L1–L2, 2 disc pairs of L2–L3, 3 disc pairs of L3–L4, and 2 disc pairs of L4–L5 were selected as queries. Half of them had osteophytes. Two of them had slight osteophytes and one was moderate and two were severe.

Twenty queries comprising of: 10 discs from pairs of cervical and lumbar vertebrae outlines, respectively, were used in the evaluation. The set of retrieved vertebrae pairs were limited to the top 20 pairs for each query generating a result set of 200 vertebrae pairs for cervical and lumbar vertebrae queries, respectively.

5.1. Experiment design

Using these data, four retrieval experiments were conducted. First, only three features (*mean and standard deviation of distance and skewness*) directly related to describing DSN were used. Second, the same three features were used but with voting consensus described in Section 4. The results of these two experiments were presented in our previous work [34]. Third, additional three features (normalized inertias) related to shape properties were used. Finally, all six features were used with feature ranking and weighting as described in Section 3 and Eq. (12) and voting consensus described in Section 4. All experiments used Procrustes distance to measure similarity between corresponding inferior and superior profiles as a complementary assistance for accurate retrieval.

For the second and fourth experiments, three or six features and voting consensus were used to generate $N + 1$ sets of retrieval results. Combining another set of retrieval results by Procrustes distance measure, these $N + 2$ sets of retrieval re-

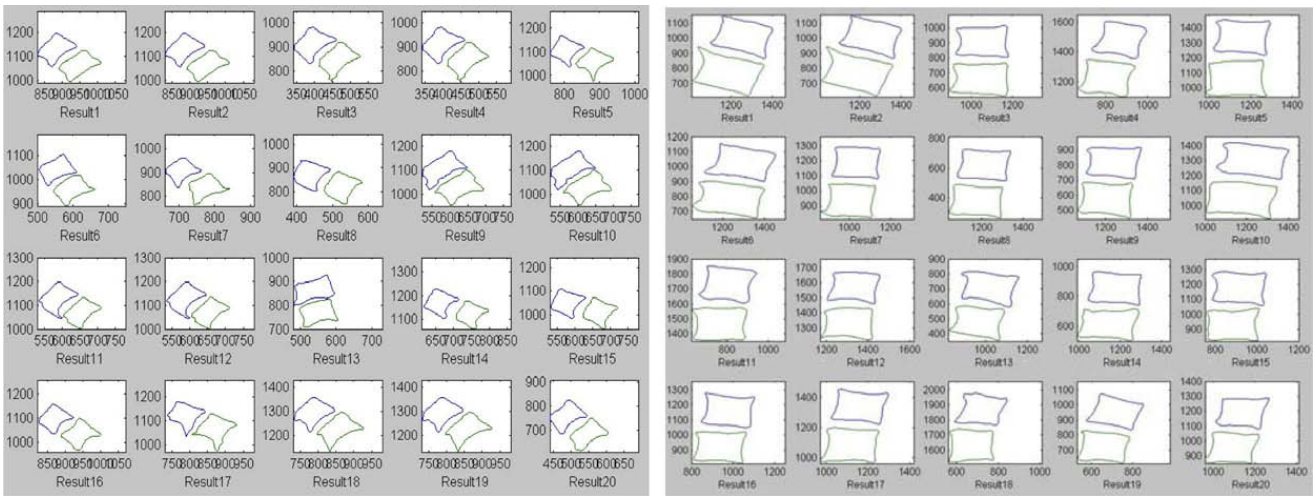

Fig. 4. Retrieval results of (a) cervical and (b) lumbar disc pairs using three features directly describe DSN properties and three features related to shape properties *without* feature ranking and weighting and voting consensus.

sults were applied to obtain retrieval sorting by voting consensus. Comparisons between the first two and the last two experiments help to understand if voting consensus improves retrieval accuracy based on DSN properties. Comparing the last two experiments helps to prove that additional shape related properties and the proposed feature ranking and weighting processes improve retrieval accuracy even further.

5.2. Retrieval results

Retrieval results using mean and standard deviation of distance and skewness with equal weights were presented in our previous work [34]. Samples of top 20 retrievals using direct retrieval with and without voting committee for cervical and lumbar disc pairs are shown to demonstrate the effect of voting consensus. In this paper, retrieval results using six features without (Fig. 4) and with (Fig. 5) feature ranking and weighting and voting consensus are shown to demonstrate the improvement of the proposed algorithm. It is noted that the first retrieval result (in the upper left corner) is the query itself and the remaining nineteen retrievals are listed in the order of their similarity to the query. The same cervical and lumbar pairs were used for both experiments.

The cervical query disk space shows the characteristics that middle section of the disk space has larger gap than the end sections. The retrieval results were reviewed by a radiologist. In Fig. 4a (cervical query), results 7, 12, 13, and14 were marked as not similar (only 16 of 20 were good matches). In Fig. 4b (lumbar query), results 5, 15, and 16 were marked as not similar (17 of 20 were good matches). Only one retrieval result was marked as not similar in Fig. 5a (result 12) and b (result 15).

From the retrieval result shown in Fig. 5a, it can be seen that most retrieved disc spaces demonstrated similar geometric shape as the query disk space. Whereas more retrieval results in Fig. 4 and results shown in [34] failed to demonstrate this shape detail. This improvement is the result of using three additional features that related to shape properties and using feature ranking and weighting to assigned different weights to all six features according to their importance. In viewing these sample retrieval results, a layperson, can verify the method appears to perform as desired. It is necessary, however, for trained eyes to evaluate its performance.

5.3. Performance evaluation

Recall and precision are commonly used as measures of image retrieval performance. They are defined in the equations below:

$$Precision = D/D + D_F \quad Recall = D/D + D_M \tag{17}$$

where D is the number of retrievals marked as good matches to the query and D_F is the number of retrievals marked as not similar to the query. D_M is the number of good matches in the database. The precision measured here is for the retrieval system as a whole, which is the combination of the similarity measure and voting consensus, rather than the matching algorithm alone. Recall was not used to evaluate the performance because each query pair must have its matches and non-matches identified in the database in order to establish the ground truth. This means for each query pair all 900 pairs must be reviewed, which is almost an impossible task for the radiologist to perform. Although there exist more complete methods to evaluate the performance of a retrieval system [35], using only "precision" measurement for our work is sufficient.

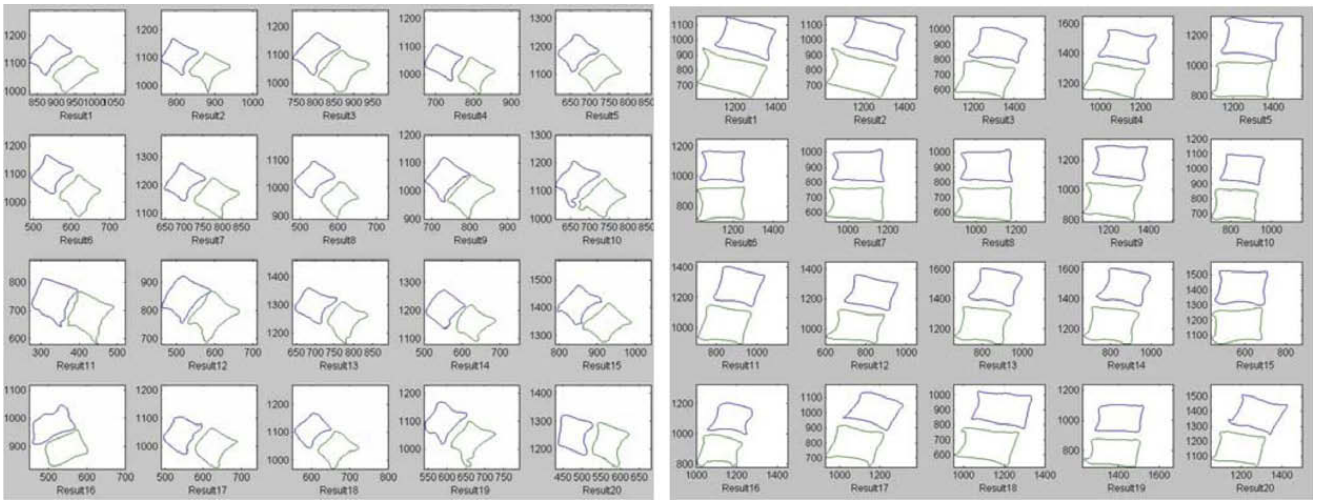

Fig. 5. Retrieval results of (a) cervical and (b) lumbar disc pairs using three features directly describe DSN properties and three features related to shape properties *with* feature ranking and weighting and voting consensus.

Table 2
Retrieval precision improvement due to three additional features, feature ranking, and voting consensus.

	Cervical	Lumbar	Precision (%)
3 Features without voting	123	185	77.00
3 Features with voting	140	189	82.25
6 Features without ranking and voting	130	190	80.00
6 Features with ranking and voting	151	190	85.25

Table 2 shows the retrieval results that were validated by a board certified radiologist specializing in diseases of the spine. Using three features without the use of voting consensus mechanism, 123 cervical pairs of 200 retrieved (61.5%) and 185 of 200 (92.5%) lumbar pairs were validated as relevant, respectively. The overall retrieval precision was 77% (308/400). In contrast, using three features with the proposed voting consensus mechanism [34], 140 of 200 (70%) retrieved cervical pairs and 189 of 200 (94.5%) retrieved lumbar pairs were validated as relevant, respectively. The overall retrieval precision of the proposed algorithm was 82.25%, which showed an improvement over 77% from direct retrieval without voting consensus [33].

When using additional three features that are related to shape properties and ranking the importance of all six features, the precision was further improved to 151 of 200 (75.5%) and 190 of 200 (95%) for cervical and lumbar disk pairs, respectively (85.25% overall). It was noted that without feature ranking and voting consensus, the retrieval precision (80.0% over all) was lower than using 3 features but with voting consensus. It was also noted that lumbar disk pairs had higher retrieval precision than cervical. This was due to lumbar pairs generally having more uniform shape and consistent DSN shape properties. Since cervical pairs vary more in shape, they pose a greater challenge for finding relevant matches thereby lowering the precision.

6. Conclusion

This paper presents a novel approach for content-based retrieval of vertebra pairs using spatial, geometrical, and shape constraints applied to inter-vertebral disc space and using both the 9-point model familiar to radiologists and bone morphometrists and the computationally meaningful 36-point vertebral shape profiles. The mean and standard deviation of disc space distances and skewness measures are used as the spatial and geometrical properties of DSN. Furthermore, features such as normalized inertias, disk space characteristics between upper and lower vertebra are extracted as similarity measures. Inter-vertebral disc shape profiles are matched using Procrustes distance as assisting features. A voting committee is constructed based on the retrieval results using six features as well as the vertebral shape profile similarity which improves the retrieval accuracy by 8.25% (77–85.25%). This is significant in applications with large image collections where queries tend to be specific to the localized pathology. It is also interesting to note that although the overall retrieval precision improved slightly compared to the earlier approach for DSN severity classification, the inclusion of shape profiles and shape properties allows the proposed method also to be used for CBIR applications.

Acknowledgement

This work research was supported by the National Library of Medicine (NLM) under Contract No. HHSN276200700335P and the intramural research funds of the Lister Hill National Center for Biomedical Communications, the National Library of Medicine (NLM), and the National Institutes of Health (NIH).

References

[1] Fact Sheet: Osteoarthritis. American College of Rheumatology, Atlanta, Georgia, 1994.
[2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.
[3] I. El-Naqa, Y. Yongyi Yang, N.P. Galatsanos, R.M. Nishikawa, M.N. Wernick, A similarity learning approach to content-based image retrieval: application to digital mammography, IEEE Transactions on Medical Imaging 23 (10) (2004) 1233–1244.
[4] J. Kim, W. Cai, D. Feng, H. Wu, A new way for multidimensional medical data management: volume of interest (VOI)-based retrieval of medical images with visual and functional features, IEEE Transactions on Information Technology in Biomedicine 10 (3) (2006) 598–607.
[5] A. Mojsilovic, J. Gomes, Semantic based categorization, browsing and retrieval in medical image databases, International Conference on Image Processing, vol. 3, 2002.
[6] H. Greenspan, A.T. Pinhas, Medical image categorization and retrieval for PACS using the GMM-KL framework, IEEE Transactions on Information Technology in Biomedicine 11 (2) (2007) 190–202.
[7] T.M. Lehmann, B.B. Wein, D. Keysers, M. Kohnen, H. Schubert, A monohierarchical multiaxial classification code for medical images in content-based retrieval, IEEE International Symposium on Biomedical Imaging (2002) 313–316.
[8] W. Liu, Q.Y. Tong, medical image retrieval using salient point detector, in: IEEE-EMBS 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, pp. 6352–6355.
[9] X.Q. Xu, D.J. Lee, S.K. Antani, L.R. Long, A spine X-ray image retrieval system using partial shape matching, IEEE Transactions on Information Technology in Biomedicine 12/1 (2008) 100–108.
[10] W. Mao, W.W. Chu, The phrase-based vector space model for automatic retrieval of free-text medical documents, Data and Knowledge Engineering 61 (1) (2007) 76–92.
[11] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications – clinical benefits and future directions, International Journal of Medical Informatics 73 (1) (2004) 1–23.

[12] L.R. Long, G.R. Thoma, Image query and indexing for digital X-rays, in: SPIE Conference on Storage and Retrieval for Image and Video Databases VII, vol. 3656, 1999, pp. 12–21.
[13] P. Chamarthy, R.J. Stanley, G. Cizek, Rodney. Long, S. Antani, G. Thoma, Image analysis techniques for characterizing disc space narrowing in cervical vertebrae interface, Computerized medical Imaging and Graphics 28 (2004) 39–50.
[14] H.G. Avad, M.S. Kamel, Cumulative voting consensus method for partitions with variable number of clusters, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (1) (2008) 160–173.
[15] T.L. Berg, D.A. Forsyth, Animal on the web, IEEE International Conference on Computer Vision and Pattern Recognition 2 (2006) 1463–1470.
[16] G. Sigletos, G. Paliouras, C. Spyropoulos, M. Hatzopoulos, Combining information extraction systems using voting and stacked generalization, Journal of Machine Learning Research 6 (2005) 1751–1782.
[17] Y. Chen, J.Z. Wang, R. Krovetz, CLUE: clustering-based retrieval of image by unsupervised learning, IEEE Transactions on Image Processing 14 (8) (2005) 1187–1201.
[18] S.H. Baloch, H. Krim, Flexible skew-symmetric shape model for shape representation, classification, and sampling, IEEE Transaction on Image Processing 16 (2) (2007).
[19] A. Srivastava, S.H. Joshi, W. Mio, X. Liu, Statistical shape analysis: clustering learning and testing, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4) (2005).
[20] A. Gersho, Asymptotically optimum block quantization, IEEE Transactions on Information Theory 25 (4) (1979) 373–380.
[21] G.A.F. Seber, Multivariate Observations, Wiley, 1984.
[22] I.L. Dryden, K.V. Mardia, Statistics Shape Analysis, first ed., Wiley, 1998. ISBN 10: 0471958166.
[23] M.L. Kherfi, D. Zhou, Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples, IEEE Transactions on Image Processing 15 (4) (2006) 1017–1030.
[24] A. Grigorova, F.G.B. De Natale, C. Daqli, T.S. Huang, Content-based image retrieval by feature adaptation and relevance feedback, IEEE Transactions on Multimedia 9 (6) (2007) 1183–1192.
[25] L. Wang, L. Khan, W. Wu, Automatic image annotation and retrieval using weighted feature selection, in: IEEE International Symposium on Multimedia Software Engineering, pp.435–442, 2004.
[26] H. Stoppiglia, G. Dreyfus, R. Dubois, Y. Oussar, Ranking a random feature for variable and feature selection, Journal of Machine Learning Research 3 (2003) 1399–1414.
[27] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (1–3) (2002) 389–422.
[28] J.G. Dy, C.E. Brodley, Feature selection for unsupervised learning, Journal of Machine Learning Research 5 (2004) 845–889.
[29] G.J. McLachlan, K.E. Basford, Mixture models: inference and applications to clustering, Dekker, New York, 1988.
[30] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of Royal Statistical Society Series B (Methodological) 39 (1) (1977) 1–38.
[31] W.M. Rand, Objective criteria for the evaluation of clustering methods, Journal of the American Statistical Association 66 (336) (1971) 846–850.
[32] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, 1967, pp. 281–297.
[33] Y.C. Chang, S.K. Antani, D.J. Lee, K. Gledhill, L.R. Long, P. Christensen, CBIR of spine X-ray images on inter-vertebral disc space and shape profiles, in: Proceedings of The 21st IEEE Symposium on Computer-Based Medical Systems (CBMS), Jyväskylä, Finland, June 17–19, 2008, pp. 224–229.
[34] D.P. Huijsmans, N. Sebe, How to complete performance graphs in content-based image retrieval: add generality and normalize scope, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2) (2005) 245–251.

Dah-Jye Lee received his B.S.E.E. from National Taiwan University of Science and Technology in 1984, M.S. and Ph.D. degrees in electrical engineering from Texas Tech University in 1987 and 1990, respectively. He also received his MBA degree from Shenandoah University, Winchester, Virginia in 1999.

He is currently a Professor in the Department of Electrical and Computer Engineering at Brigham Young University. He worked in the machine vision industry for eleven years prior to joining BYU in 2001. His research work focuses on Medical informatics and imaging, shape-based pattern recognition, hardware implementation of real-time 3-D vision algorithms and machine vision applications.

He is a senior member of IEEE and a member of SPIE. He has actively served the research community as a paper and proposal reviewer and conference organizer.

Sameer Antani received his B.E. (Computer) from University of Pune in 1994, M.E. and Ph.D. degrees in Computer Science and Engineering from the Pennsylvania State University in 1998 and 2001, respectively.

He is currently a Staff Scientist with the Lister Hill National Center for Biomedical Communications an intramural R&D division of the National Library of Medicine which is an institute within the US National Institutes of Health. His research interests are in data management of and retrieval from, large biomedical multimedia archives. His research includes content-based indexing, and retrieval of biomedical images, combining image and text retrieval, and next-generation documents that are enriched with interconnections to data sets and multimedia content.

He is a member of IEEE and IEEE Computer Society. He serves on the steering committee for IEEE Symposium for Computer Based Medical Systems (CBMS). He is a reviewer for several journals including various IEEE transactions.

Yuchou Chang was born in Hunan, China in 1980. He received his B.S. degree in automatic control department from Northwestern Polytechnical University, Xi'An, China, in 2003 and M.S. degree in institute of image processing and pattern recognition from Shanghai Jiao Tong University, Shanghai, China, in 2006. He worked in the Robotic Vision Laboratory in the Electrical and Computer Engineering Department at Brigham Young University as a research assistant from September 2006 to December 2008.

His research interest includes machine learning-assisted multimedia analysis, image segmentation, image and video semantic content description, content-based multimedia indexing and retrieval. He is an IEEE student member.

Kent Gledhill was born and raised in Utah. He attended undergraduate and medical school at the University of Utah in Salt Lake City, graduating in 1992. He completed his Radiology residency and Neuroradiology fellowship at the University of New Mexico in Albuquerque. He currently practices diagnostic and neuroradiology in Utah County. He serves on the Board of Directors of the Central Utah Clinic in Provo and has been the medical staff president and a board member at Timpanogos Regional Hospital in Orem. He lives in Provo, Utah with his wife and four children.

L. Rodney Long received his B.A. and M.A. degrees in mathematics from the University of Texas in 1971 and 1976, respectively, and M.A. degree in applied mathematics from the University of Maryland in 1987.

Since 1990 he has been an Electronics Engineer in the Communications Engineering Branch of the National Library of Medicine. He previously worked for 14 years in industry as a programmer and engineer. His research work is concerned with Content-Based Image Retrieval, image processing, and image databases for biomedical applications.

He has been a member of IEEE since 1986 and has served as co-chair of the IEEE International Symposium of Computer-Based Medical Systems.

Paul Christensen is currently an undergraduate student in the Computer Science Department at Brigham Young University. He is expecting to receive his B.S. degree in December 2009. He served a two-year church mission for The Church of Jesus Christ of Latter-day Saints in Edinburgh, Scotland from 2004 to 2006. He worked as a research assistant in the Robotic Vision Lab at BYU from 2006 to 2008 and is currently an intern at Intel in Hillsboro, OR.